

KEEPING IT REAL

BART DEYGERS & KOEN VAN GORP
CENTRUM VOOR TAAL & ONDERWIJS, KU LEUVEN

ALTE 45TH MEETING & CONFERENCE, LONDON, NOVEMBER 2014

INNOVATION

ALTE 45th Meeting and Conference, London, 19 November 2014

ALTE's 45th bi-annual Meeting and Conference will be hosted by Trinity College London and will take place in London, 19 - 21 November, 2014.

The first two days of meetings will be for representatives of ALTE Members and Institutional Affiliates only, and the final day, Friday 21st November, will be an Open Conference Day for all those with an interest in language testing. The theme of the conference day will be *Innovation in Language Testing*.

The conference will be preceded by a two-day Auditor Orientation and Training Session, and a two-day Testing Course on 17-18 November, and will be followed by a one-day Foundation Course on 22nd November.

INNOVATION?

For decades, entrepreneurs and digital gurus of various repute have referred to this era, in a breathlessness bordering on proselytizing, as the age of innovation. But *Weird Stuff* is a reminder of another, unexpected truth about innovation: It is, by necessity, inextricably linked with failure. The path to any success is lined with disasters. Most of the products that do make it out of the lab fail spectacularly once they hit the market. Even successful products

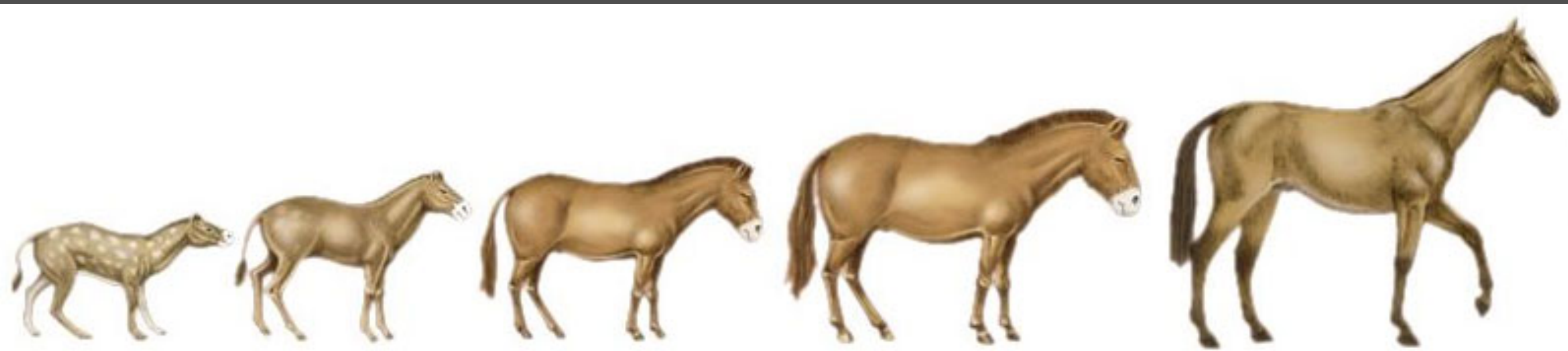
INNOVATION?



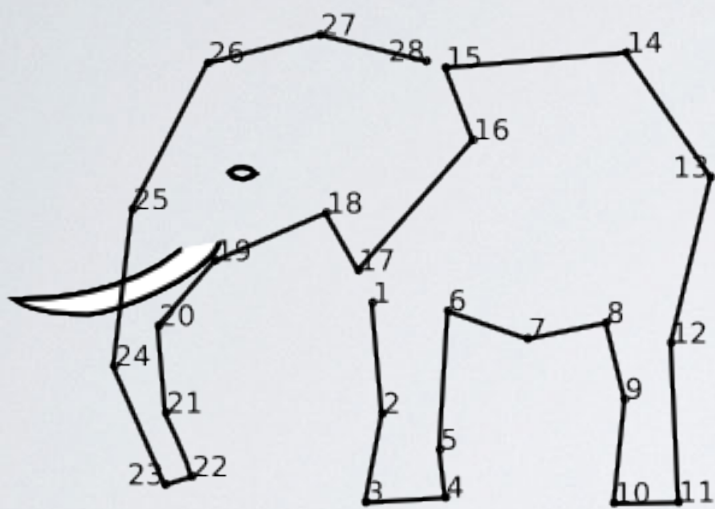
Q: Age is no guarantee for efficiency.

Bond: And youth is no guarantee for innovation.

INNOVATION OR EVOLUTION?

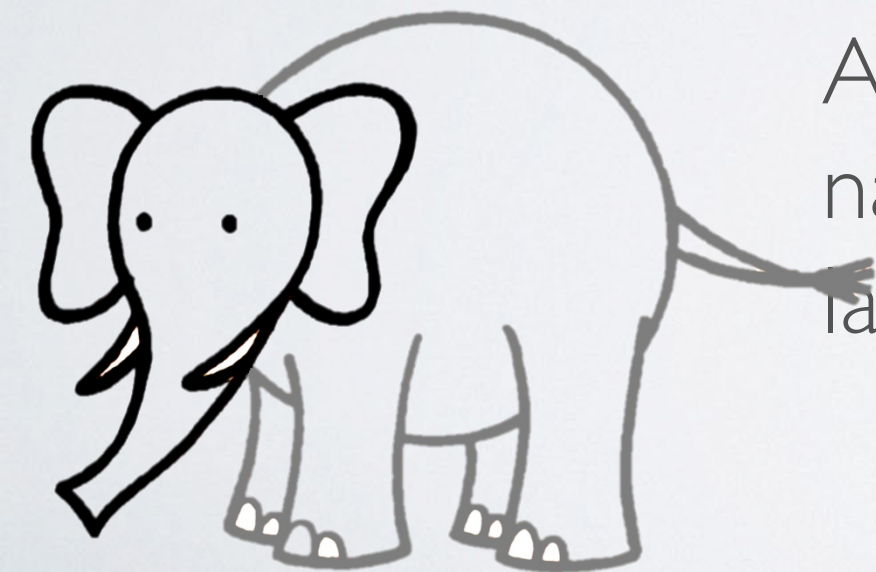


TWO TRADITIONS



A test is a measurement tool which, given enough data, is able to accurately gauge one's language ability

Cf. De Jong



A test is a performance-based argumentative narrative about what test takers can do with language

Cf. Norris

THE HOLY GRAIL OF TESTING?



ARE WE THERE?

“TBLA takes the task itself as the fundamental unit of analysis, motivating item selection, test instrument construction, and the rating of task performance”

(Long & Norris, 2000)



Realistic, multifaceted & challenging activities
Specific content
Meaningful context
Personally relevant and motivating

(Cumming, 2013; Norris, 2014)



Task dependency: tasks confound the measurement
TBLA involves hybrid genres that are ill-defined
Task fulfillment requires a certain threshold level
Problems with inferencing
Murky measurement

(Bachman, 2002; Weir, 2005, Cumming 2013)

NOT QUITE

Inferencing & generalizing

What does a performance tell us about real-world ability?

NOT QUITE

Inferencing & generalizing

Scoring validity

How consistent are the raters? How valid are the criteria?

NOT QUITE

Inferencing & generalizing

Scoring validity

Context validity

How representative is this task for the TLU context?

NOT QUITE

Inferencing & generalizing

Scoring validity

Context validity

Theory-based validity

Does the task measure the construct at hand?

NOT QUITE

Inferencing & generalizing

Scoring validity

Context validity

Theory-based validity

Test-taker characteristics

Task difficulty resides in the interaction with the learner

(Bachman, 2002; Weir, 2005; Norris & Ortega, 2009; Révész, 2009; Gan, 2012)

A CENTURY OF SOLUTIONS

Write an Essay on one of the following topics (2 hours)

The effect of political movements upon 19th century England

English Pre-Raphaelitism

The Indian Mutiny

The development of local self-government

Matthew Arnold

(Cambridge English Test, 1920s)

A CENTURY OF SOLUTIONS

Write an English Essay on one of the following subjects (3 hours)

Noise

Castles in Spain

The sculptor

Tolerance

Imitation is the sincerest form of flattery

Our mechanized civilization

(Cambridge English Test, 1960s)

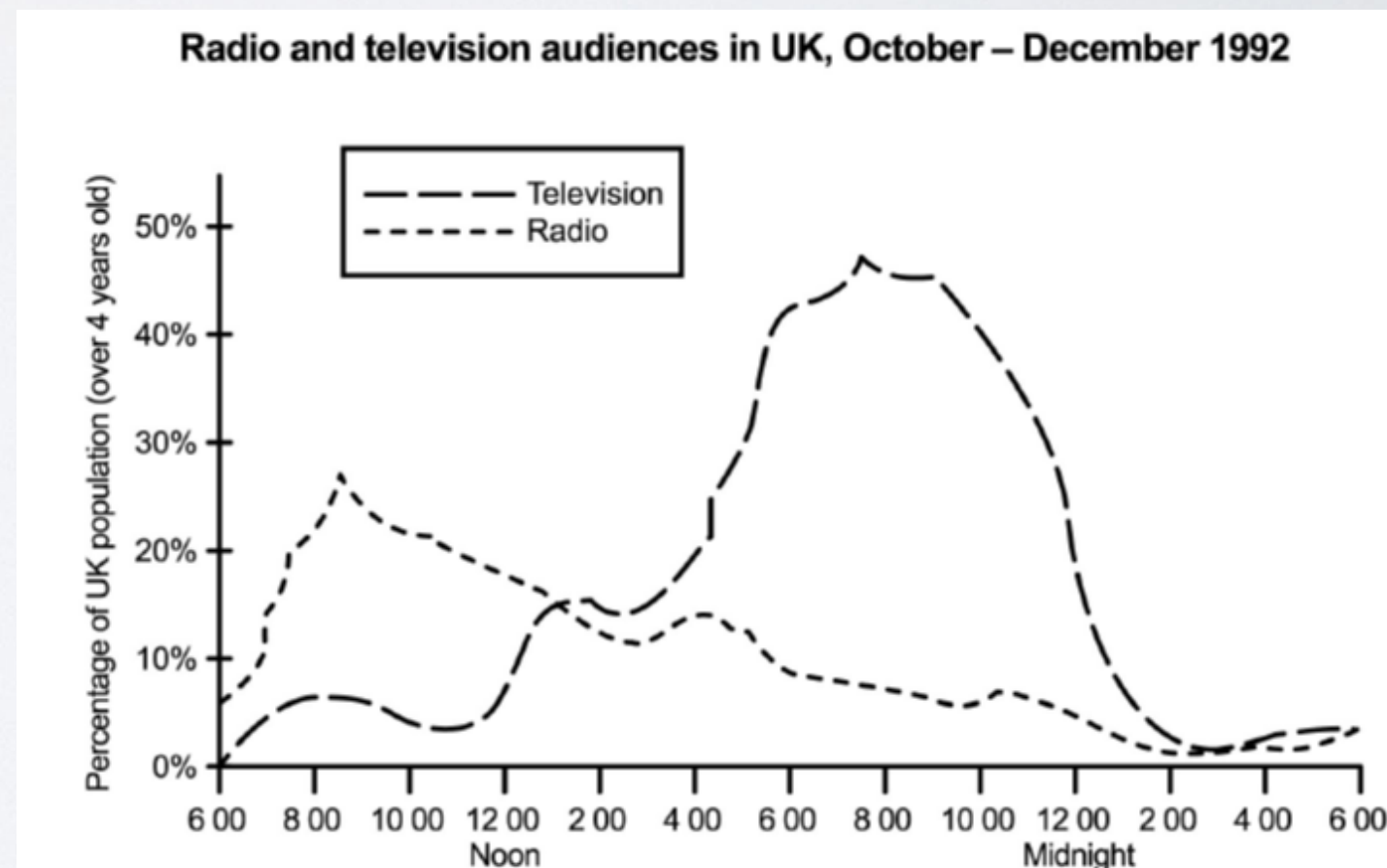
A CENTURY OF SOLUTIONS

Writing task

You should spend 20 minutes on this task

The graph below shows radio and television audiences throughout the day in 1992. Summarize the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words



(Cambridge English Test, 2010s)

TASKS IN PRACTICE



DEVELOPING TASKS

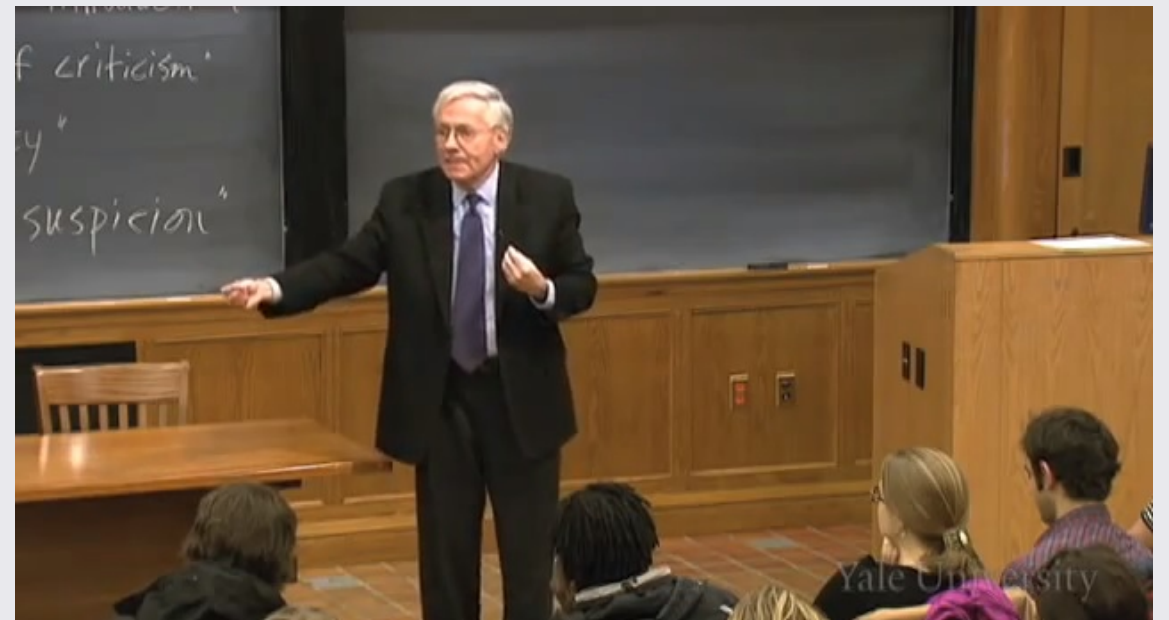
1. Specify intended uses for the assessment
2. Select and analyze key target tasks/features from needs analysis
3. Design tests and items: emphasize authentic performances
4. Determine real-world criteria for rating task performance
5. Pilot-test and revise instruments and procedures
6. Evaluate validity in terms of intended uses, and washback

A NOTE TAKING TASK

Calcium

A student of medicine, you are attending a course on nutritional sciences. Since some of your friends cannot attend class, you are writing a summary for them.

Listen to the lecture and write a consistent 150-word summary. If you write less, your text might receive a lower score. Pay attention to style, lexis and grammar.



MEET MOHAMMEDREZA

Age	23
Nationality	Iran
L1	Persian
L2 instruction	6 months
Study	Civil Engineering
Test score	Fail

AUTHENTICITY

Test



In the previous class we discussed vitamins in our food. Today we will deal with calcium. Calcium in our food. I hope that you have brought your syllabus because at the end of class I will give you an assignment. But first I will talk about the importance of calcium in our food. On the television you often see commercials advertising milk. They tell us to drink more milk. In order to stay healthy, we should drink some 4 glasses of milk, the ads say...

AUTHENTICITY

Class

↓
This afternoon we are going to discuss a specific kind of improper integral, called the Fourier transformation, which fulfills a very important role in the analysis of signals. The action of this transformation goes as follows: observe this function of t . It's not a function of x but of t because I will be interpreting the variable as a variable of time. My f of t will be a time signal. However weird it may sound but for the first time in this course we will use a function which can obtain values in \mathbb{C} ...



AUTHENTICITY

Medicijnen kopen via de internet, gewoontje en zelfs...
dadelijk zijn, daarom ben ik tegen het kopen van
medicijnen via het internet. Medicijnen moet
gekeurd zijn bij de dokter of bij een apotheker...

Als je via de internet een medicijn koopt...
ben je onzeker wat je eigenlijk koopt. Je hebt
geen kennis van wat er achter zit en wat de
samenstelling is. Als je medicijn van de dokter
of van de apotheker dan weet je zeker dat professionele
medicijnen hebben gemaakt en ook
weet zeker dat ze de medicijnen hebben gekocht.

Online is het heel makkelijk om te kopen
en hierdoor loop je grote risico's om genees-
middelen te kopen van mensen of websites die
niet zeker de kwaliteit zijn of echte websites die
niet bestaan is van het kopen van medicijnen...

Als je online geneesmiddelen wilt kopen loop je
en groot risico. Het is beter om gewoon een
recept van je dokter te ontvangen en medicijnen
kopen bij de apotheken. Dan ben je zeker weet...

← Expectation

AUTHENTICITY

Actual notes →

Als de functie f stukgewijze Continu is in $[R, P]$
 $\forall R, P$ en bovendien absoluut riemann-integreerbaar is in \mathbb{R}
 $\Rightarrow f(\omega)$ cont op \mathbb{R} en $f(\omega) \rightarrow 0$ als $\omega \rightarrow \pm\infty$

als f riemann int \mathbb{R}

~~$\int_{-\infty}^{+\infty} |f(t)| dt$~~ Continu

$$\left| \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \right| \leq \int_{-\infty}^{+\infty} |f(t)| e^{-i\omega t} dt$$

door $\frac{1}{\omega}$

beperkt

$$\int_0^b \frac{\sqrt{1-x^2}}{x} dx$$

INFERENCES, HUH?

Possible construct-relevant variance if academic listening tests exclude visual cues.

(Lynch, 2001; Wagner, 2008; 2010)

Theoretical listening validity is not enhanced through integrated writing tasks.

(Weir, 2005; Wolfersberger, 2013)

Input influences output

(Cubilo & Winke, 2013; Yu, 2013)

Task – candidate interaction

(Rost, 2001; Bachman, 2002; Davies, 2008; Norris & Ortega, 2009; Révész, 2009; Gan, 2012)

SCOPE & CONCESSIONS

Broad scope

Scoring validity	==> Less indigenous criteria
Context validity	==> Broader TLU
Theory-based validity	==> What construct?
Test-taker characteristics	==> Less clean profile



Realistic, multifaceted & challenging activities
Specific content
Meaningful context
Personally relevant and motivating

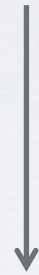
TBLA OR NOT?

Task-based assessment vs Task-supported assessment

(Ellis, 2003; Norris, personal communication)

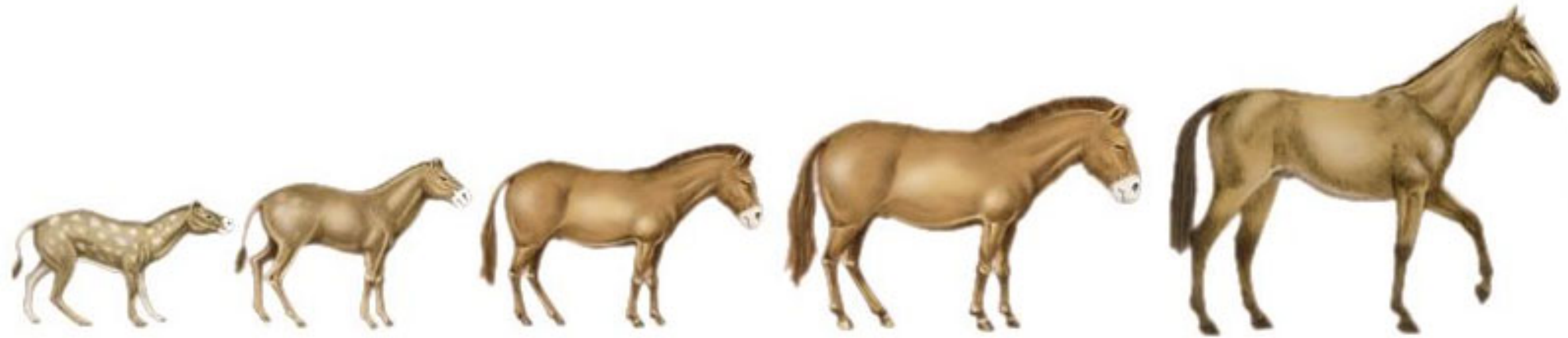


Task used to elicit specific communicative functions within specific contexts



Task as tool to elicit generically rated language performance

ALTERNATIVES



“We need to go as far as we can”

(Hyland, 2002)



TBLA IN CLASS

In TBLA, tasks are essential pedagogic constructs that “drive” classroom activity

(Samuda & Bygate, 2008)

“Ultimately, language learning does not depend on a few ‘super moments’, but on thousands of interactional moments ...”

(Van den Branden, 2010: 296)

Or - lots of tasks:

Assessment tasks as learning tasks (Carless, 2007)

Learning tasks as assessment tasks

TBLA IN CLASS



The disappearance of Mr. Baldido (reading)

TBLA IN CLASS



The disappearance of Mr. Baldido (reading)

Design your own hero (reading, speaking)

Look for clues (reading, speaking)

Search the castle (listening)

Solve riddles (reading, speaking)

Write a letter (writing)

The final battle (listening)

Escape punishment at trial (speaking)

Report to the local newspaper (writing)

THE CHALLENGE

TBLA has to provide “frameworks for tracking and interpreting important aspects of learner development over time”.

(Norris, 2009: 587)

Consequently, for teaching & (formative) assessment purposes, tasks should be conceptualized as a set of characteristics, rather than holistic entities.

(Bachman, 2002)

MEETING THE CHALLENGE

A coherent analytic framework for task-based classroom assessment:

- contains a set of characteristics/task specifications

- allows for interpretations of task performance (strategies)

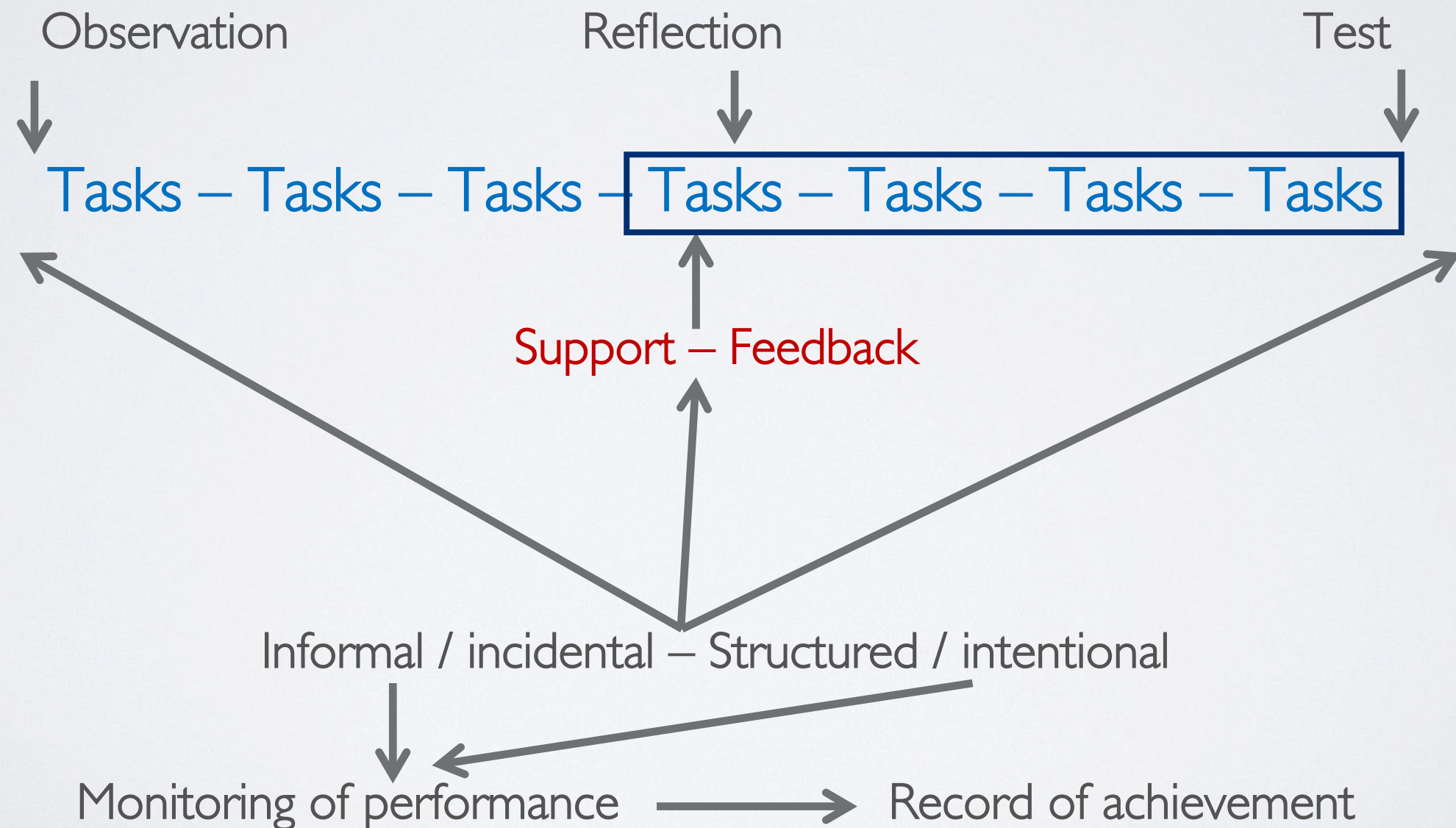
- enables tracking of learner development

- is integrated in syllabus design, curriculum development, curricular thinking

(Byrnes, 2013)

TBLA MODEL

Teacher – Other students – Self



COMPONENTS OF TBLA

Observation of task performance (incidental)

to provide teacher support, feedback if necessary ('dynamic assessment')

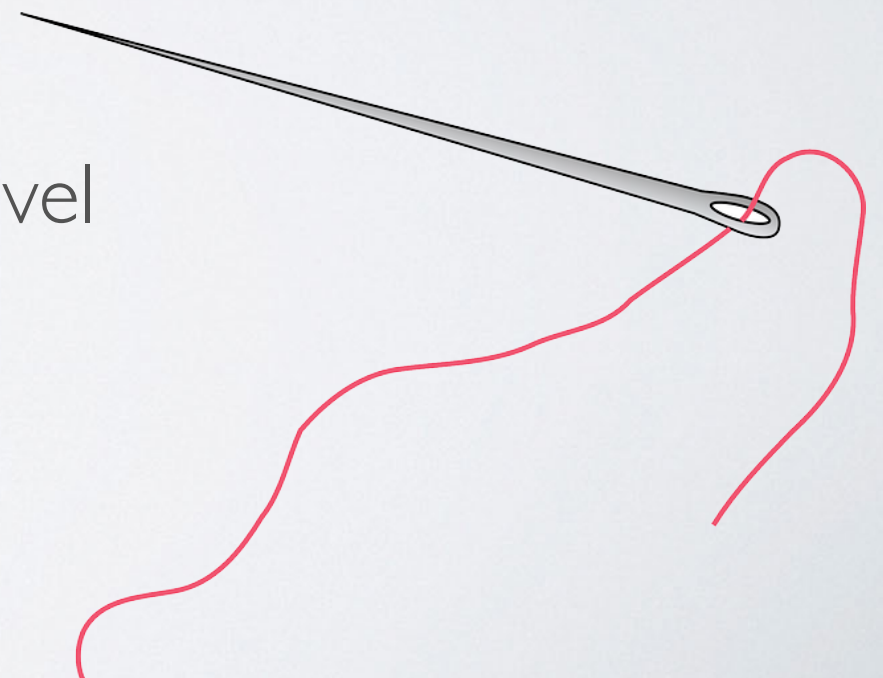
Systematic observation and analysis of task performance and task outcome
of all or "targeted" individual students (intentional)

Learner reflection and portfolio

of task performance and language proficiency level

Task-based tests

for summative use



A TASK SPECIFICATION FRAMEWORK

Table 35.1 Task specification framework for the reading task “Which way out?” (5th grade) in *TotemTaal* (Berben et al., 2008a). Adapted with permission from the authors

<i>Goals</i>							
Parameters▶	Skill	Level of processing	Text-type	Public	World	Function	Attainment goal
Settings▶	Reading	Evaluating	Informative texts	(Un) known peers	Orientation (scientific description)	Inform	3.4

Enables task sequencing,
Monitors task complexity, and
Tracks learning opportunities
Exceeds mere task fulfillment

= backbone of a task-based curriculum and a TBLA framework

IN-CLASS TBLA



Real-world activities?

Concessions?



KEEPING TASKS OUTSIDE

Workplace Dutch:

Learn Dutch (A2/B1) on the job

Preparatory year of Dutch:

B1 entrance test, then do it

- ✓ Scoring validity
- ✓ Context validity
- ✓ Theory-based validity
- ✓ Test-taker characteristics



Realistic, multifaceted & challenging
Specific content
Meaningful context
Personally relevant and motivating

5. ARE WE PIONEERS?



5. ARE WE PIONEERS?



(Not really)

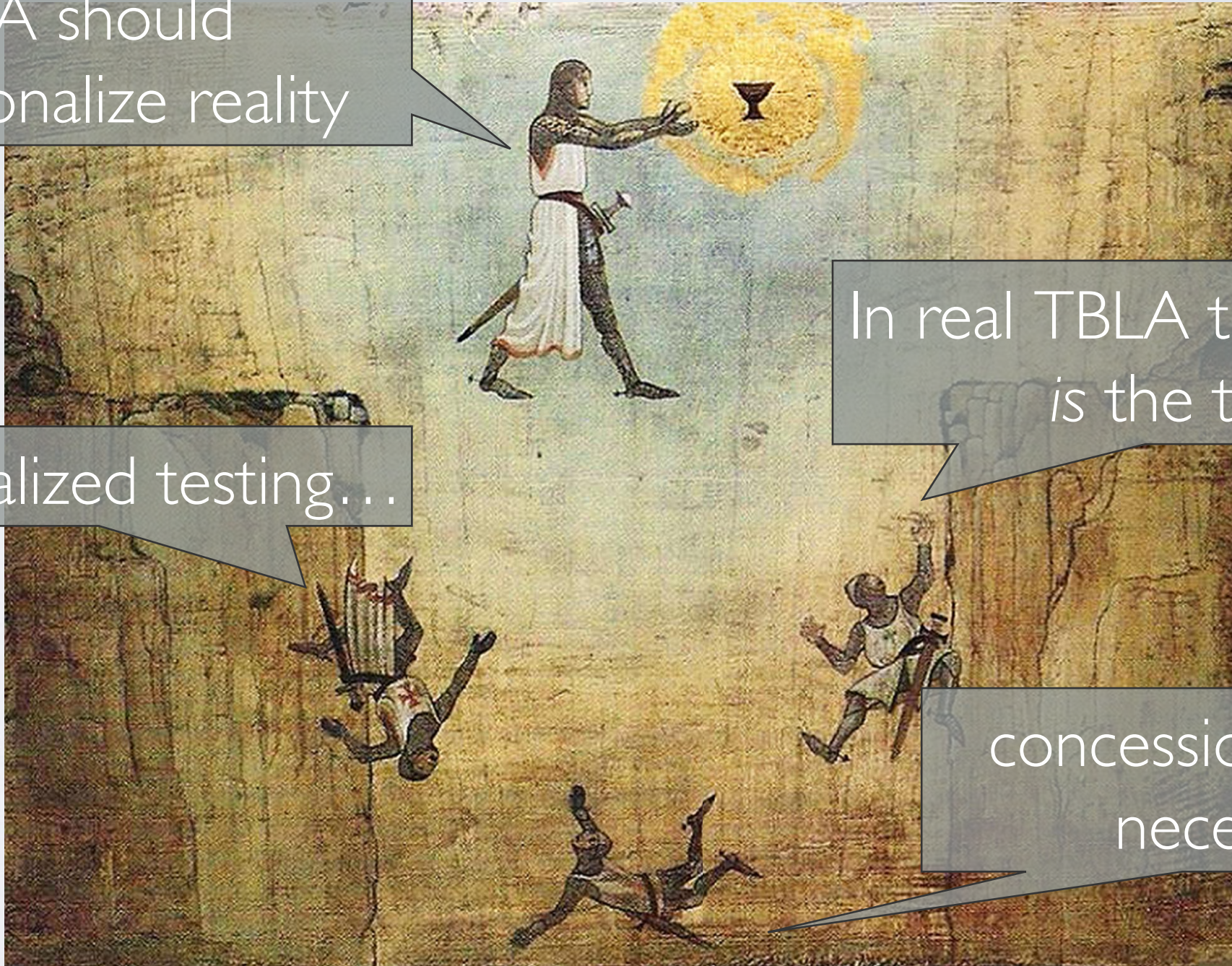
THE HOLY GRAIL OF TESTING?

TBLA should operationalize reality

but in centralized testing...

In real TBLA the construct is the task...

concessions are a necessity



Thank you!

bart.deygers@arts.kuleuven.be

koen.vangorp@arts.kuleuven.be

